Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

European Journal of Mechanics B/Fluids 25 (2006) 172–191

A conditional sampling method based on fuzzy clustering for the analysis of large-scale dynamics in turbulent flows

G. Usera [a,b], A. Vernet [a], J. Pallares [a], J.A. Ferré [a,*]

[a] Departament D'Enginyeria Mecànica, Universitat Rovira i Virgili, Av. Països Catalans 26, 43007 Tarragona, Spain
[b] Instituto de Mecánica de los Fluidos, Universidad de la República, J.H. Reissig 565, 11300, Montevideo, Uruguay

Received 24 April 2004; accepted 26 June 2005

Abstract

A conditional sampling technique, based on fuzzy clustering, is used to educe the organization of the secondary flow motions observed in the large-eddy simulation of a turbulent square channel flow. The data analysed are the multi-valued time series obtained from sampling the secondary velocity components at a fixed cross-section of the channel, over consecutive time steps. The mean values of the secondary flow motion velocities are one order of magnitude lower than their r.m.s values. The purpose of the conditional sampling scheme used is to replace the picture of the secondary flow motions provided by the unconditional time mean with several ensemble averages. In this way the whole variability of the instantaneous data can be split in two parts: one for the difference between the observed ensemble averages, and the other for the variability within each ensemble. Unlike other conditional sampling schemes which sort only part of the data into one or more families depending on an externally fixed condition, the fuzzy clustering approach used here first determines the optimum number of families or clusters and then classifies all the recorded time steps. The results show that the local turbulence intensities of the ensemble averages obtained from fuzzy clustering can be reduced by one order of magnitude. In addition, the classification of all the time steps into several clusters or families enables the large scale dynamics of the educed structures to be analysed.

Keywords: Fuzzy clustering; Conditional sampling; Turbulent flow

---

1. Introduction

The increasing availability of full field measurement techniques with good time resolution, such as time resolved Particle Image Velocimetry (PIV), and of higher order numerical simulation methods for turbulent flows, such as Large Eddy Simulations (LES) and Direct Numerical Simulations (DNS), provides the fluid dynamic researcher with databases that use extended time series to give more details about the flow. These databases tend to get larger and larger and specialised methods need to be developed if they are to be analysed efficiently. This paper analyses the extended time series of velocity fields in planar cross sections of high turbulence intensity flows. Turbulent flow in a square duct has been selected to illustrate the proposed method of analysis.

---

[*] Corresponding author. Tel.: +34977559602; fax: +34977559691.
E-mail address: josep.a.ferre@urv.net (J.A. Ferré).

0997-7546/$ – see front matter © 2005 Elsevier SAS. All rights reserved.
doi:10.1016/j.euromechflu.2005.06.002

Turbulent flow in a duct is the prototypical case of turbulent flow along a stream-wise corner, where secondary flows of Prandtl's second kind arise. Despite yielding a weak mean cross-stream velocity field, these secondary flows have a significant effect on flow properties of great practical interest (e.g. wall shear stress, and heat and mass transfer), as has been previously recognised by other authors (e.g. [1–3]). Some of the models used to solve the Reynolds-averaged Navier–Stokes equations are unable to reproduce secondary flows properly, so either LES or DNS are preferred. After more than a decade of numerical experiments on turbulent flow in square ducts, both DNS (e.g. [4,1,2]) and LES (e.g. [5–7]), have been established as valid techniques for gaining further insight into the dynamics of the flow. Once these simulations have been validated against experimental data, they provide a full and detailed picture of the flow, which can then be analysed by the appropriate techniques so that the flow dynamics can be understood better.

Proper orthogonal decomposition (POD) is one such technique that involves computing a subset of the eigenvectors and eigenvalues of the correlation tensor, in decreasing order of the magnitude of the eigenvalues. The eigenvectors with the largest eigenvalues can provide insight into the underlying large-scale structure of the flow in a square channel [8]. Kopp et al. [9] considered using these leading eigenvectors as starting templates for pattern recognition to search for coherent structures. In addition, the projection of the original data set onto the subset of eigenvectors makes it possible to reduce the dimensionality of the data set with a loss of information that is limited to the finer, and thus less energetic, scales of the flow. This is how POD is applied here. A summary of the applications of the POD technique has been given by Berkooz et al. [10]. More recently, Johansson and George [11] and Garnard et al. [12] further considered its application in the study of turbulent flows.

Another orthogonal decomposition technique has been proposed by Farge et al. [13] as an alternative to POD. The coherent vortex simulation (CVS) is a non-linear wavelet-based decomposition technique that adapts the number and selection of resolved modes, and keeps only the strongest wavelet coefficients at each time step. It thus rivals POD, which is a strategy with a fixed number of resolved modes, in terms of compression efficiency.

Pattern recognition (PR) is another set of techniques used to identify flow structures that govern the dynamics of the flow [14,15]. It involves cross-correlating initial templates, or prototypical structures, with the instantaneous events contained in the time series. Those events whose correlation factors exceed a user-defined threshold are accepted to create the ensemble average for the educed structure. However, usually only a small number of the instantaneous events contained in the time series are involved in the pattern classification process. For instance in Kopp et al. [9], the group of classified instantaneous events accounts for only 35–38% of the flow. On the other hand, fuzzy clustering (FC) has recently proved to be a useful technique that can classify all the instantaneous frames into a specified number of conditionally averaged or coherent subsets [16]. This technique is applied here to identify the flow structures that are hidden within the usual unconditional averaging procedure.

The purpose of this paper is to present a novel conditional sampling method based on a fuzzy clustering technique and its application to the analysis of secondary flow structures of the turbulent flow in a square duct. To overcome the limitations of unconditional time averaging, a set of conditionally sampled averages, or cluster averages, are obtained with an improved signal-to-noise ratio. The time evolution of the secondary flow organisation is also analysed, which provides the most probable sequence in time of the conditionally sampled averages. This two-step procedure (conditional sampling and time ordering) finally yields a simplified view of the flow dynamics of the secondary flow motions in the square channel. Following this introduction, the data set used and the analysis techniques will be described in detail in Sections 2 and 3. Section 4 presents some results of the analysis of the cross-stream velocity components, while Section 5 describes the time evolution analysis of the flow. Finally, the conclusions are summarised in Section 6.

2. Working data set

The data analysed in this paper was obtained from the large eddy simulation of the fully developed turbulent flow in a square duct at $Re_\tau = u_\tau h/\nu = 300$, which corresponds to $Re_h = u_b h/\nu = 4500$ [17]. A sketch of the computational domain and the nomenclature for the velocity components is shown in Fig. 1. The domain has an $h \times h$ square cross section and its length is $2\pi h$. In the cross section, the computational grid contained 66×66 points, non-uniformly spaced. However, the grid was uniform in the streamwise direction, with 66 nodes and periodic boundary conditions. The computational time step was $\Delta t = 5 \times 10^{-3} h/u_\tau \approx 8 \times 10^{-2} h/u_b$ and the data series was sampled at every time step.

The initial working data set was a time series of 1400 instantaneous, three-component, consecutive velocity fields, recorded at a planar cross section of the square duct in a 66×66 grid. Therefore, there were $66 \times 66 = 4356$ velocity

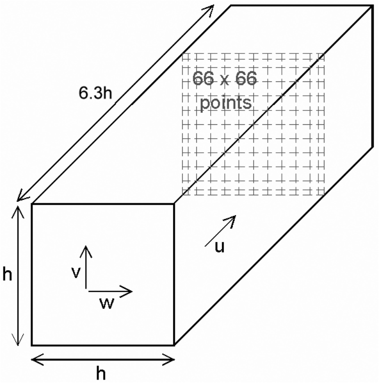

Fig. 1. Sketch of the computational domain and velocity components.

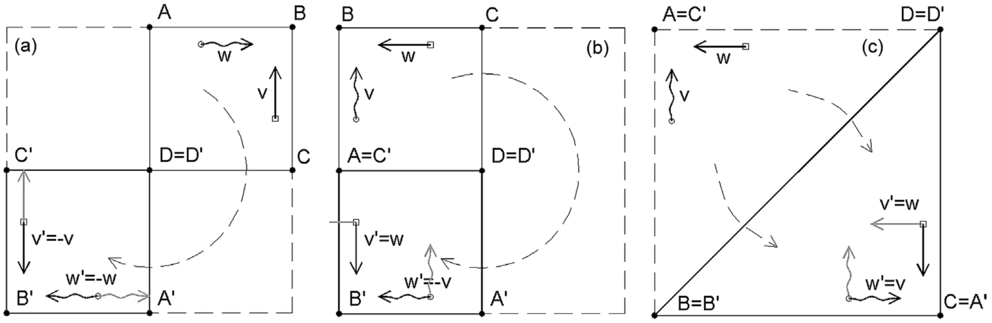

Fig. 2. (a) Transformation of the upper right quadrant into the lower left quadrant. (b) Transformation of the upper left quadrant into the lower left quadrant. (c) Transformation of the upper octant into the lower octant of the lower left quadrant.

vectors per time step. However, since the statistical properties of the flow are expected to be the same in any given octant, the full square duct data was used to generate a time series of $8 \times 1400 = 11200$ instantaneous velocity fields containing the data of the single octants of the square duct. Other authors have already proposed increasing the number of flow samples through this argument of symmetry (see [18]). Each frame corresponded to an octant of the square section, with a reduced number of only 561 velocity vectors per time step. Data from each octant was kept in a continuous sequence of 1400 time steps, in such a way that the time discontinuities of the whole data set were reduced to only 7 cuts in a series of 11200 time steps. The different octants were rotated and flipped and (v, w) were also transformed as needed, to make them all consistent. Only the u component of the velocity remained unchanged throughout these transformations (see Fig. 2). No spurious correlations were introduced using this procedure because the length of each continuous time series segment (1400 time steps) is two orders of magnitude larger than the integral time scale of the flow $T_\tau = 14.9\Delta t$, which is dominated by the streamwise velocity correlation, as can be observed in Fig. 3(c).

Fig. 3(a) presents the vector plot of the mean cross-stream velocity field and a contour plot of the mean stream-wise component. Indirectly, Fig. 3(a) also shows the non-uniform computational grid. Fig. 3(b) shows the mean stream-wise component of the wall shear stress along the wall (normalised by the mean wall shear stress). These results are consistent with those presented by other authors [4,18,1,2]. An extensive comparison between the results of the present study and previous DNS and experimental measurements can be found in Pallares and Davidson [17], and in further detail in Pallares and Davidson [19].

However, the averaged picture of the flow in Fig. 3(a) gives misleading information about the dominant characteristics of the instantaneous velocity fields, since the mean values displayed in Figs. 3(a) and 3(b) are up to one order of magnitude weaker than their respective r.m.s. values. Therefore, in this case one cannot expect to find a single instantaneous velocity field or shear stress profile that looks like the conventional time mean and these mean velocity fields say very little about the true dynamics of the secondary flow motions.

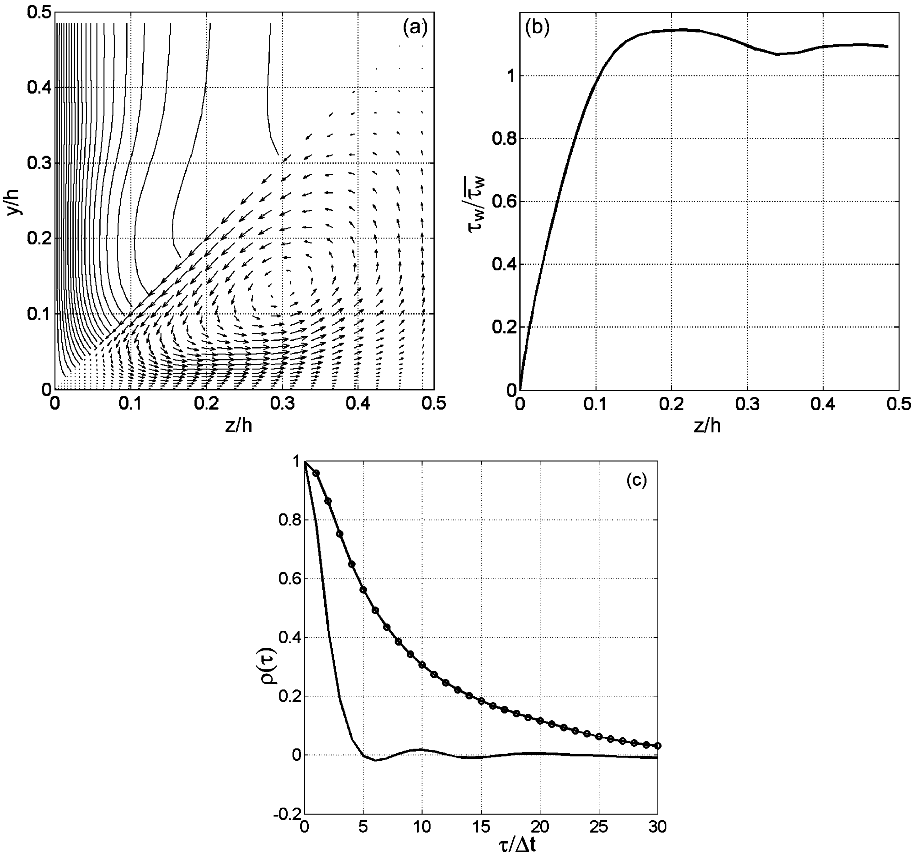

Fig. 3. (a) Octant mean velocity fields: u component contour lines (top) and (v, w) component secondary flow vector field (bottom). (b) Stream-wise component of mean wall shear stress along bottom half of the wall. (c) Auto-correlation coefficient $\rho(\tau)$ for the cross stream velocity field (—) and streamwise velocity (—○—), averaged over the octant section.

Fig. 3(c) displays the time auto-correlation of both the (v, w) cross stream velocity field and the streamwise velocity component. The correlation coefficient was computed at each grid point and then averaged over the octant. It can be observed that the first one goes to zero for a delay of five time steps, and that the time-memory of the (v, w) velocity field beyond that point fluctuates slightly below and above zero, but with very low values. As noted previously, the memory of the flow is dominated by the u component, which shows a much smoother auto-correlation decay. This relatively sudden decay of the (v, w) auto-correlation can be explained by the comparative analysis of several consecutive time steps. While the individual eddies retain their identity for time sequences larger than those expected from the information in Fig. 3(c), they move continuously within the square duct, mainly from the centre towards the walls. This continuous change in the relative position of the secondary eddies within the duct quickly generates a near zero correlation, as a result of zones of positive correlation pooling with zones of negative correlation. The auto-correlation of the streamwise velocity decays in a time delay of about 30 time steps which corresponds to a downstream distance of $x/h = 2.3$, using the relation $u_b \approx 15.0u_\tau$. This result is in agreement with the values reported in Fig. 3 of Gavrilakis [4].

3. Analysis technique

The main objective of the analysis technique applied here is to substitute the conventional time mean by several ensemble averages obtained from a reasonable conditional sampling procedure in such a way that the data shows reduced variability around these conditionally averaged ensembles. The procedure that enables us to do this is fuzzy clustering. However, although FC can be performed with the raw data set, a POD analysis of the raw data was first

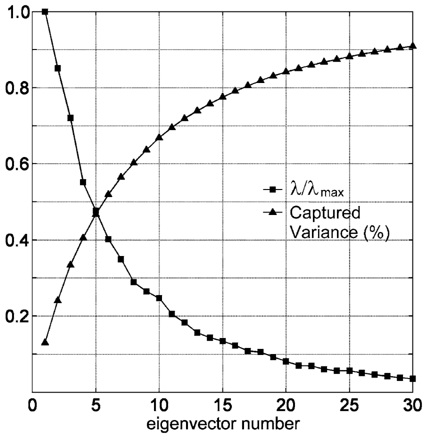

Fig. 4. POD normalised eigenvalues (—■—) and cumulative captured variance (—▲—).

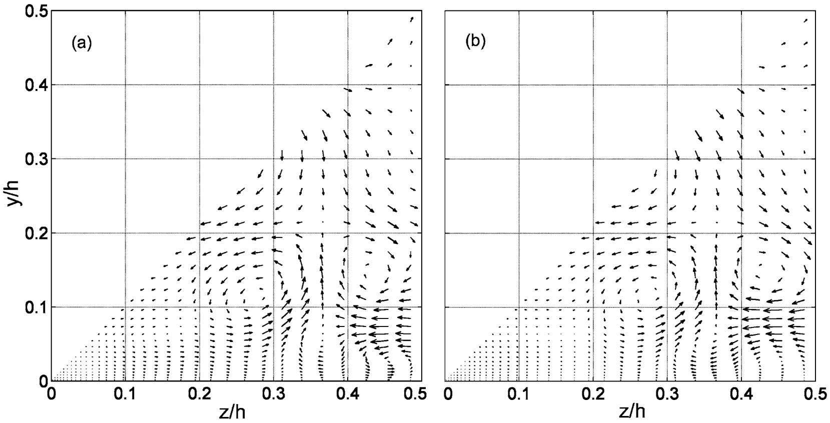

Fig. 5. (a) Instantaneous secondary flow velocity field. (b) POD projected field.

applied in order to compress the volume of data handled by the FC procedure, and to focus the FC classification on the larger scales of the secondary flow. POD was preferred to other alternative decomposition techniques, such as the wavelet transform based CVS technique [13], mainly because it is simpler. The classification operations that are performed afterwards by FC are also more straightforward.

The cross-stream velocity field within the octant consists of two velocity components sampled at 561 points. Therefore each time step can be thought of as a single point in a 1122-dimensional space. POD showed that the first 30 eigenvectors captured more than 90% of the total variance of the set, with the next eigenvector accounting for less than 5% of the total variance. These results are summarised in Fig. 4, which plots both the cumulative captured variance and the normalised eigenvalues for the first 30 eigenvectors. Gavrilakis [18], working with a reduced region near the wall from a DNS at a Reynolds number similar to the value used here, needed 17 modes to retain 90% of the turbulent kinetic energy. However, Reichert et al. [8] used 100 modes to retain 95% of the variance or turbulent kinetic energy of the DNS data computed at twice the Reynolds number used here.

Subsequently, each instantaneous velocity field is represented by its projection onto these eigenvectors, the data set being effectively reduced from its original 1122-dimensionality to a 30-dimensional space. This involves a reduction of more than one order of magnitude in the size of the data set, making it far more manageable and allowing the use of analysis techniques that still consider the data set as a whole. If the data set were not to be reduced, some operations on the whole set might be inefficient. This problem is greater if numerical simulations are performed at larger Reynolds numbers and/or if direct numerical simulations (DNS) are considered instead of LES, because in both cases the data sets get larger and larger. Still, the nature of POD ensures not only that the loss of information is small,

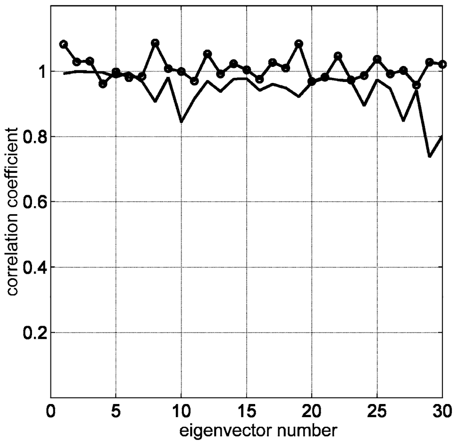

Fig. 6. Correlation between POD performed on independent halves of the data set. Corresponding eigenvector correlation (—-) and ratio between eigenvalues (—○—).

but also that it is constrained to the finer scales of the flow. Fig. 5(a) shows an instantaneous cross-stream velocity flow field, while Fig. 5(b) presents the reconstruction of this field using the first 30 eigenvectors. Note that there are only minor differences between Figs. 5(a) and 5(b): for example, the smearing out of a small vortex near the wall at $z/h = 0.45$ and $y/h = 0.05$.

While analysis of the POD eigenvectors can yield some insight into the flow structures, POD by itself cannot provide any information about the time history of the flow, since the cross-correlation matrix keeps no information on the time ordering of the velocity frames. However, this information can be accessed by analysing the time history of the projection components of the velocity frames onto the POD eigenvectors, as will be explained in Section 5. It should be pointed out that while the POD eigenvectors constitute an optimal, orthogonal base for decomposing each velocity frame into a linear combination of eigenvectors, they are not the best templates to be used to classify instantaneous velocity frames. This is due to their hierarchical and spectral nature. In other words, instantaneous velocity frames, or their given ensemble averages, should not be expected, in general, to reassemble any single POD eigenvector, but a suitable linear combination of eigenvectors. This is especially true for eigenvectors associated to low eigenvalues, because they tend to contribute to the finer scales.

To corroborate that the data set is large enough for our analysis, we performed an auxiliary POD analysis on two independent halves of the data set. The resulting sets of eigenvectors and eigenvalues were correlated to obtain the results presented in Fig. 6, which shows that both sets are nearly completely correlated, within 10% for most modes, with slight deterioration towards the higher modes associated with the finer scales.

Fig. 7 shows the first five eigenvectors plus the twentieth eigenvector to illustrate this. As can be observed in Fig. 4, the first five eigenvectors capture nearly 50% of the variance of the data while their individual normalised values are still higher than 0.5. The twentieth eigenvector was selected since it is the first one that displays a normalised value lower than 0.1 (that is to say, one order of magnitude weaker). While the first few eigenvectors might resemble some of the ensemble averages that will be presented later, it is clear that, as the eigenvector sequence advances, they increasingly reproduce the finer structures, as expected from the hierarchical and orthogonal nature of the POD decomposition. While these properties of POD are desirable in a decomposition method, they are neither necessary nor desirable in a classification method. Something else that has to be taken into account in the analysis of Fig. 7 is that these velocity fields can be fully sign-reversed without losing their condition of being the true eigenvectors of this POD analysis. Therefore, POD cannot distinguish between secondary flow motions with positive (into the plane) or negative (out of the plane) vorticity.

As stated above, the objective is to obtain a set of ensemble averages that substitute the conventional time mean, so that the data exhibit reduced variability, or fluctuating kinetic energy, around this set. Having been projected onto the first 30 eigenvectors, each instantaneous velocity field of the time series analysed (indexed as $j = 1, 2, \ldots, n$, with $n = 11\,200$) is just a vector X_j in a 30-D space, with an unconditional mean $\overline{X}$. If the data is classified in $i = 1, 2, \ldots, c$

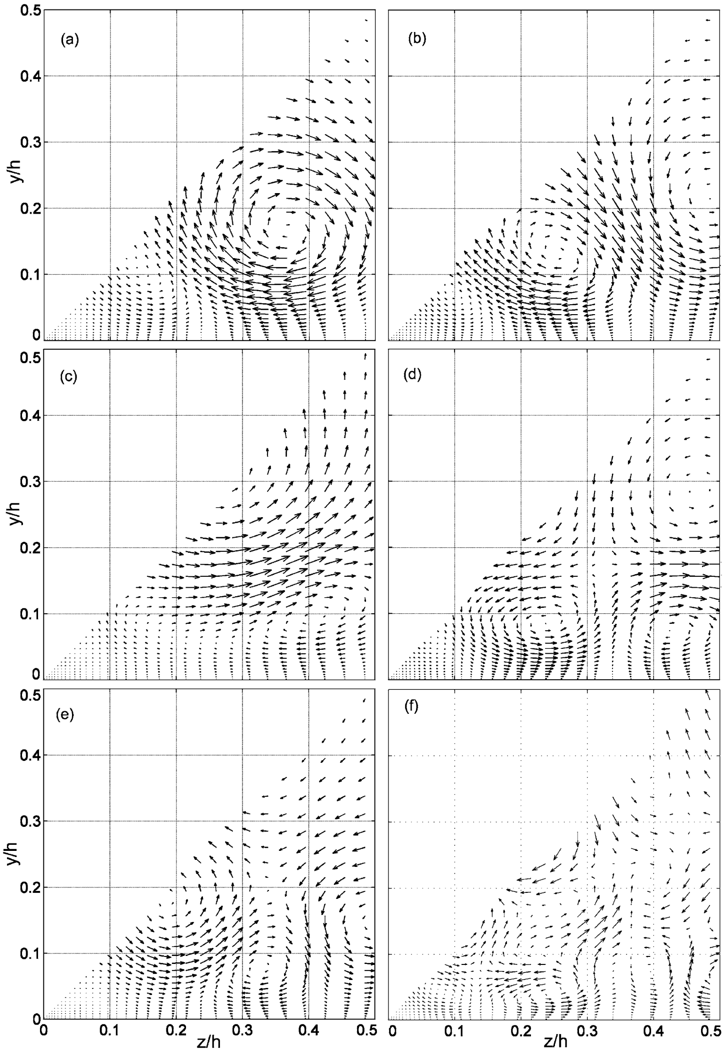

Fig. 7. Selected POD eigenvectors. (a)–(e) Eigenvectors #1–5. (f) Eigenvector #20.

different families (or clusters) each of them with their own mean V_i, then the whole variability of the data (the term on the left-hand side of Eq. (1)) is distributed among the new c clusters according to

$$\sum_{j=1}^{n} d^2(X_j, \overline{X}) = \sum_{i=1}^{c}\left(\sum_{k=1}^{n_i} d^2(X_{ki}, V_i)\right) + \sum_{i=1}^{c} n_i \cdot d^2(V_i, \overline{X}) \tag{1}$$

where the initial X_j $(j = 1, 2 \ldots, n)$ vectors on the left-hand side term have been sorted into $i = 1, 2 \ldots, c$ clusters in the terms on the right-hand side, and are now denoted as X_{ki} $(k = 1, 2 \ldots, n_i$ vectors in cluster $i)$. Also $d^2(\cdot,\cdot)$ is used here to note the Euclidean distance between two points

$$d^2(X_i, X_j) = \|X_i - X_j\|^2. \tag{2}$$

The first term on the right-hand side of Eq. (1) is the sum over all the clusters ($\sum$ for $i = 1, 2, \ldots, c$) of the variability within each cluster ($\sum$ for $k = 1, 2, \ldots, n_i$), while the second term is the contribution from the c cluster means. A classification that would suit the purpose of our analysis would be one that maximizes this second term on the right-hand side of Eq. (1), but minimises the first one, that is, the unexplained variability. Note that what we have been generically calling "variability" is, in fact, the kinetic energy of the fluctuating (v', w') velocity field since

$$d^2(X_j, \overline{X}) = \sum_{\text{whole field}} v'^2 + \sum_{\text{whole field}} w'^2. \tag{3}$$

Therefore, we can now say that our initial purpose of "reducing the variability" exhibited by the data around the conventional time-mean effectively means "reducing the kinetic energy" of the fluctuating velocity field not explained by the c cluster means.

The aim of some algorithms used for fuzzy clustering, and in particular Bezdek's c-Means Clustering algorithm [20], is closely related to this purpose, since it tries to minimize an objective function that is very similar to the first term on the right-hand side of Eq. (1). Strictly, for a prescribed number of clusters c, and fuzziness index m $(m > 1)$, the c-Means Clustering algorithm tries to minimize an objective function J_m, defined as:

$$J_m = \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^m d^2(X_j, V_i) \tag{4}$$

where X_j $(j = 1, 2, \ldots, n)$ is the data set to be clustered and $d^2(X_{j,} V_i)$ is a measure of the distance between the vectors X_j and the cluster centre V_i, taken here as the Euclidean distance, and μ_{ij} is the fuzzy membership function that verifies the condition:

$$\sum_{i=1}^{c} \mu_{ij} = 1. \tag{5}$$

For any partition, the cluster centroids are defined from:

$$V_i = \sum_{j=1}^{n} (\mu_{ij})^m X_j \bigg/ \sum_{j=1}^{n} (\mu_{ij})^m. \tag{6}$$

Also, the fuzzy membership function can be computed from a specified set of centroids V_i from

$$\mu_{ij} = \left(\frac{1}{d^2(X_j, V_i)}\right)^{1/(m-1)} \bigg/ \sum_{i=1}^{c} \left(\frac{1}{d^2(X_j, V_i)}\right)^{1/(m-1)} \tag{7}$$

J_m is minimized by iterating Eqs. (6) and (7) from a starting guess of either μ_{ij} or V_i until no further improvement in J_m (Eq. (4)) is observed.

The sum of squared distances in Eq. (4) will be equivalent to the first term on the right-hand side of Eq. (1) only if $\mu_{ij} = 1$ for the n_i vectors belonging to cluster i, and zero otherwise. This condition is verified approximately since, due to the quotient expression in Eq. (7), the membership function μ_{ij} of a vector X_j is near zero for all the clusters except for the one that is closest to this vector. Thus, during the iteration of Eqs. (6) and (7), the membership function can take any value in the range $[0, 1]$, although these values are usually closer to either 0 or 1 after the iteration converges. Once the procedure has finished, the membership function is binarised (for any data point X_j $\mu_{ij} = 1$ only for one cluster V_i and zero otherwise), so that each vector X_j is assigned to the cluster closest to it, which corresponds to the maximum value of $\mu_{ij} = 1$ for that particular X_j. Accordingly, the last clusters obtained are just conventional ensemble averages of the instantaneous (v, w) velocity fields that have been classified into the same family or cluster.

Still, the c-Means Clustering algorithm by itself gives no information about the number of clusters to be used in partitioning the set. In order to determine the optimum number of clusters to be used for each data set, as well as the value of the fuzziness index m, we applied the validity criterion for fuzzy clustering, proposed by Xie and Beni [21]. For any given partitioning, regardless of the algorithm used to determine it, a compactness and separation validity function S, can be defined as:

$$S = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}^2 d^2(X_i, \overline{X}_j)}{n \cdot \min_{ij} d^2(X_i, \overline{X}_j)}. \tag{8}$$

The validity function S can be regarded as the ratio of the measure of the compactness of the clusters over the minimum separation among clusters. Optimum values of c and m are those for which S reaches an absolute minimum value.

The last step in the description of the fuzzy clustering algorithm used to analyse the data is to briefly discuss the initialisation of the iterative solution of Eqs. (6) and (7). Some authors simply recommend a random initialisation of μ_{ij}. To check the stability of the minimum of J_m (Eq. (4)) obtained in this way, we compared several hundreds of different random initialisations of μ_{ij}. The results obtained confirm that a random initialisation of μ_{ij} always leads to the same minimum value of J_m. However, we preferred an approach that took into account the fact that the raw data had previously been pre-processed with POD. Either the strongest eigenvectors or the ensemble averages of the vectors contained in the first few quadrants (n vectors define 2^n quadrants) were used as the initial cluster centroids, That is, we used either "pure" eigenvectors or their linear compounds. In both cases the results obtained from the iterative solution of Eqs. (6) and (7) were indistinguishable, and we finally decided to use the linear compounds as our standard starting strategy.

4. Secondary flow field analysis

The application of the fuzzy clustering validity criterion (Eq. (8)) yielded optimum values of $m = 1.12$ and $c = 17$ for the classification of the cross stream velocity field. As noted above, this criterion for optimal partitioning means that there is no need to arbitrarily specify the number of cluster ensembles to be constructed. Instead, the value of this parameter results from the intrinsic characteristics of the data set and the application of the validity criterion.

When the 11200 time steps of the octant data set were clustered with these parameters, about the same number of instantaneous frames were classified into each of the 17 clusters, giving a mean number of 660 frames in each of them (or about 5.9%, see Fig. 8). Figs. 9(b), 9(c) and 9(d) present only three cluster mean (v, w) velocity fields out of the seventeen obtained, together with the unconditional mean field (Fig. 9(a)), which has been added for purposes of comparison. All the vector fields plotted in the lower octant of the figures are given the same scale so that the relative weakness of the unconditional mean (v, w) velocity field can be assessed. For the sake of clarity, the unconditional mean field is symmetrically re-plotted on a larger scale in the upper octant of Fig. 9(a). Cluster mean velocity fields are up to one order of magnitude stronger than the local unconditional time mean. The gray scale background in Fig. 9 represents the ratio between the local energy of the fluctuating (v', w') field (Ec. 3) and the respective local mean field ($\overline{X}$ in Ec. 1). Thus, this ratio is the local turbulence intensity of the cluster ensemble. In the darker areas the high relative intensity of fluctuations reaches ratios that are above 20, while in the brighter areas this ratio can be below 1. The mean ratio, averaged over the octant section, for the unconditional time-mean (Fig. 9(a)) is about 40, while for any of the three clusters presented in Figs. 9(b), 9(c) and 9(d) it is below 2. The darkest regions in Fig. 9(a) are the result of the local unconditional time mean approaching zero faster than the fluctuating (v', w') field.

Several of the features in Fig. 9 merit some comment. First, the unconditional mean (Fig. 9(a)) is relatively weaker than any of the conditionally sampled or cluster mean averages (Figs. 9(b), 9(c), 9(d)). Second, the increase in the signal-to-noise ratio displayed by the cluster means relative to the unconditional time mean. But most strikingly, the

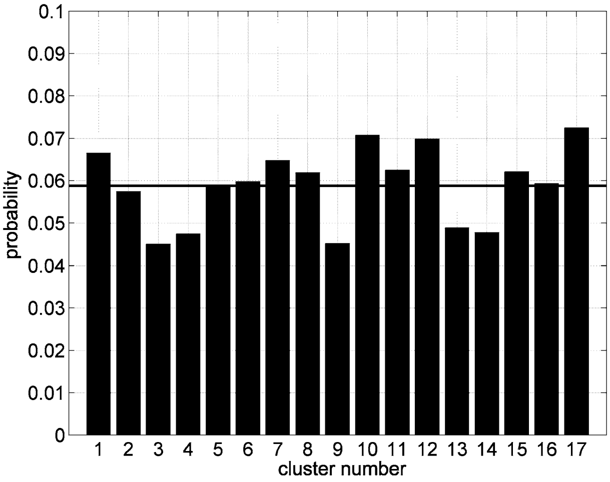

Fig. 8. Probability of occurrence of each cluster. Thick line indicates the mean probability (1/17) of equally probable clusters.

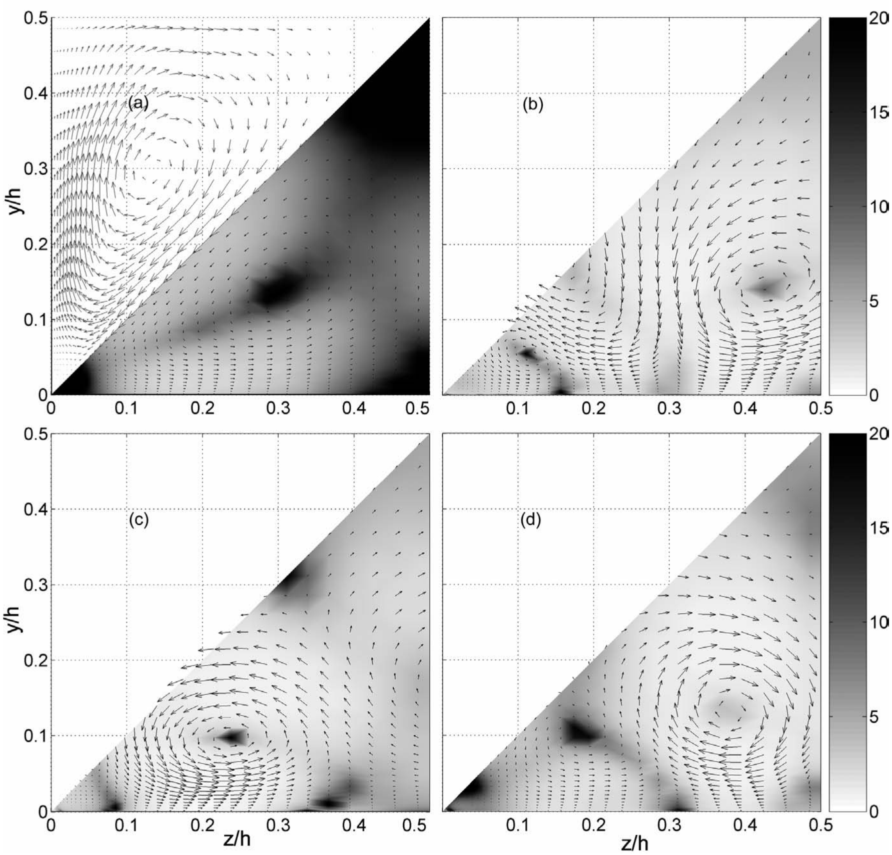

Fig. 9. Secondary flow field with local fluctuations intensity overlay. (a) Unconditional time mean. (b) (c) and (d) Selected clusters. Velocity vectors in the lower octants are plotted to the same scale in all frames. For purposes of comparison, in the upper octant of frame (a) the mean velocity field has been (symmetrically) replotted on a larger scale.

velocity fields of these true ensemble averages are very different from the unconditional mean which, in comparison, becomes meaningless. In fact the clusters in Figs. 9(b), 9(c) and 9(d) have been selected because they exhibit large-scale circulation with the same sign as the mean velocity field (Fig. 9(c)), with the opposite sign (Fig. 9(d)), or simultaneous zones with both signs (Fig. 9(b)).

Fig. 10 shows some instantaneous velocity fields selected from the clusters in Figs. 9(b), 9(c) and 9(d). While other instantaneous fields within each class show varying levels of similarity to the corresponding cluster mean, Fig. 10 emphasises the fact that, despite being ensemble averages, the cluster means do represent physically existing flow realisations.

The distribution of the kinetic energy among clusters, announced in Eq. (1), is presented in Fig. 11, where we have plotted the seventeen terms on the right-hand side of the equation divided by the left-hand side term: that is,

$$1=\sum_{i=1}^{c}\left(\frac{\sum_{k=1}^{n_i} d^2(X_{ki},V_i)}{\sum_{j=1}^{n} d^2(X_j,\overline{X})}+\frac{n_i\cdot d^2(V_i,\overline{X})}{\sum_{j=1}^{n} d^2(X_j,\overline{X})}\right). \tag{9}$$

For the sake of comparison, a horizontal line is plotted at 1/17 (i.e. 5.9%), the expected mean contribution of each cluster to the total sum of squares. Another line is also plotted at about 2.1%, which corresponds to the mean explained variability of the data; that is, the second term on the right-hand side of Eq. (9). The difference between these two values, about 3.8%, is the mean contribution from the still unexplained variability of the data; that is, the first term on the right-hand side of Eq. (9). The relative weight of the explained kinetic energy to the total kinetic energy is only moderate, about 36% ($1.8/5.9 \approx 0.36$) if it is computed as the ratio of mean squared velocities, or 59% ($0.36/0.5 \approx 0.59$) computed as the ratio of the square root of mean squared velocities. However, it is still much higher

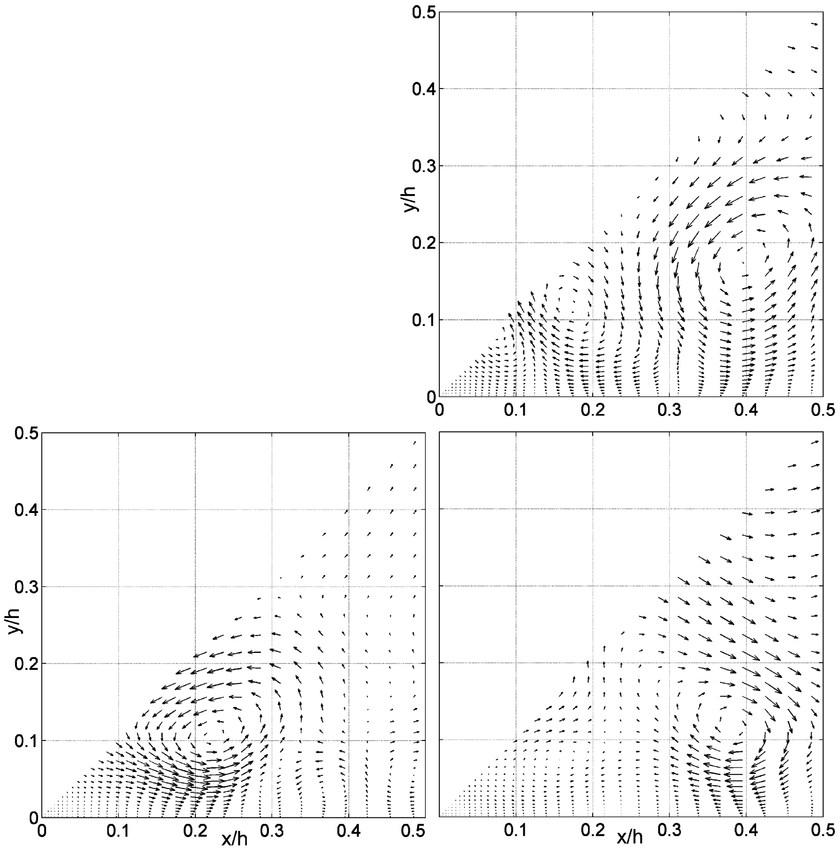

Fig. 10. Selected instantaneous velocity frames from those contributing to the cluster averages depicted in Figs. 9(b), 9(c) and 9(d).

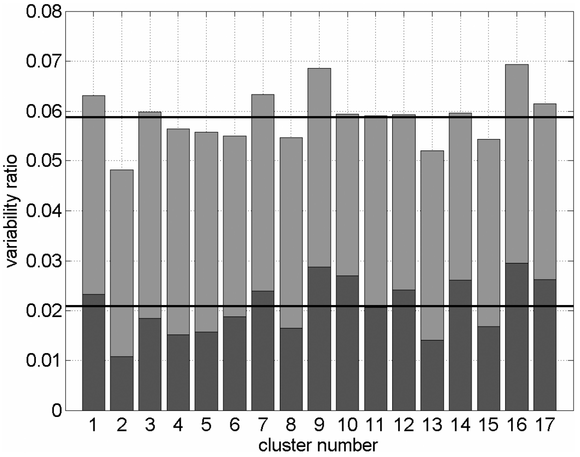

Fig. 11. Explained (dark gray, lower bars) and unexplained (light gray, upper bars) kinetic energy distribution among clusters.

than it was for the unconditional mean, where the kinetic energy of the fluctuating field was an order of magnitude above the kinetic energy of the mean field (for example, see Fig. 9).

In order to explore this subject further, we analysed how the kinetic energy content of the cluster mean velocity fields is distributed along the eigenvector spectrum. The results are presented in Fig. 12, where we compare the r.m.s. of each POD component of the whole data set and the r.m.s. of the POD components of the seventeen clusters. While

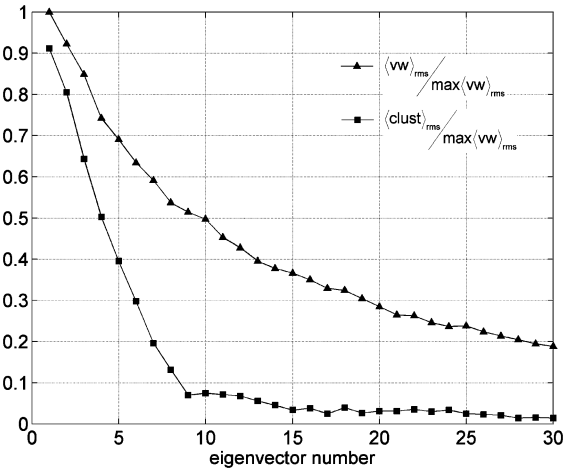

Fig. 12. Rapidly decaying energy content of cluster means along the eigenvector spectrum.

the energy content of the raw data slowly decays along the POD spectrum, the energy content of the cluster means is almost constrained to the first 10 eigenvectors, as can be expected for any ensemble average, where finer scales have been smeared out by the averaging procedure. Thus, the energy content associated to the remaining 20 eigenvectors corresponds to finer scale details and is out of the reach of the computed cluster means. This energy content can be as much as 31% of the total fluctuating energy in the cross-stream flow (see Fig. 4). Thus, from the remaining 69% of the energy associated to large scale motions, 36% has been captured by the computed cluster means or ensemble averages, while 33% is still attributed to the random character of the turbulent secondary motions.

Incidentally, Fig. 12 also justifies the choice, made in Section 3, of keeping just the first 30 eigen modes of the POD. Since the obtained cluster means only have significant projections onto the first 10 modes, the results of the present analysis would not be affected by the number of modes retained, as long as this number is comfortably larger than 10. The discarded modes would have no impact on the obtained cluster ensemble averages.

Below the fuzzy clustering (FC) procedure is compared with the pattern recognition (PR) and proper orthogonal (POD) techniques. For this purpose, PR was applied to the data set using the first few POD eigenvectors as initial templates for the PR analysis, as proposed in Kopp et al. [9]. The cross correlation threshold level for accepting individual elements was tuned to obtain ensemble averages consisting of a number of elements similar to those obtained by the FC technique. The resulting threshold level was 0.76. Fig. 13 displays the four ensemble averages obtained using the first four POD eigenvectors already presented in Fig. 7 as initial templates.

The ensemble averages in Fig. 13(a) and (b) are similar to their initial POD templates, while in cases (c) and (d) they are considerably different. As stated above, this behaviour is a result of the spectral nature of the POD decomposition. As we proceed higher into the POD modes, the POD eigenvectors should not be expected to resemble true ensemble averages of the flow. In other words, POD decomposes the flow into a spectral domain of orthogonal modes, while FC and PR decompose the flow into distinct ensemble averages within the spatial domain.

The circulation patterns in Fig. 13, obtained by PR, are similar to those of Fig. 9(b), (c) and (d), which were obtained by FC. Thus it seems that both PR and FC detect compatible sets of the ensemble averages which characterize the flow. However, PR requires starting templates to be provided while FC does not. FC also provides the means for obtaining a full set of disjoint ensemble averages that satisfy a given criterion of compactness and separation. Because PR determines the ensemble averages independently it is not possible, in principle, to enforce such conditions of compactness and separation. Notice that in case (a) and case (c) of Fig. 13 the ensemble averages are quite similar, even though the starting templates were quite different. Furthermore, the PR ensembles are not required to be disjoint and might share some instantaneous elements, and all the instantaneous velocity fields are not required to participate in the classification process. These properties of the fuzzy clustering process also suit the purpose of the section below, which analyses the temporal organisation of the data, since FC is able to classify all the instantaneous velocity fields in one family or another.

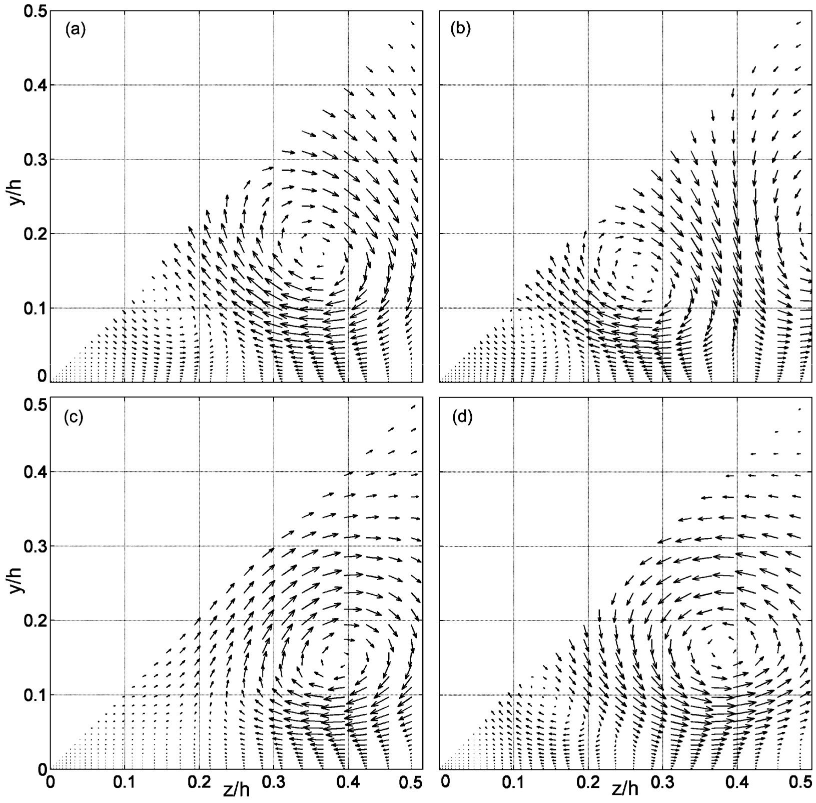

Fig. 13. Flow patterns obtained from pattern recognition using the first four POD eigenvectors from Fig. 7 as starting templates.

5. Time ordering of the clusters

All the results presented in the previous section are independent of the actual order of the time steps contained in the time series that has been analysed. That is, if they were to be in a random order, the clusters in Fig. 9, or the kinetic energy distribution analysed in Fig. 11, would be the same. Incidentally, we should also point out that the POD analysis that was reported in Figs. 4–7 is also insensitive to any random permutation in the time steps of our time series data, since the spatial correlation matrix on which POD is constructed is independent of the time sequence of the data.

To extract the data's temporal information, we will start by analysing how the instantaneous (v, w) data fields are classified into clusters. Suppose that each one of the 17 clusters obtained are labelled with a capital letter $(A, B, \ldots, Q)$. Since all the individual velocity fields are classified as belonging to one and only one of the clusters, a capital letter can also be assigned to each time step. Therefore the time evolution of the secondary motions can be viewed as a sequence like:

$$\cdots KKKPPPPBJJJHHOFFEECCCI\cdots$$

where each letter represents the label of one time step.

At the beginning of this series it can be seen that for three successive time steps the instantaneous velocity fields are classified into cluster K, then into cluster P for four time steps, and so on. Successive time steps belonging to the same cluster is known as a "tie". Some ties are very short, like the B that appears between the P's and the J's, but others are 2, 3, 4 or even more time steps long. Fig. 14 shows the tie length distribution. The mean tie length is 2.6. The tie length distribution is consistent with the fast decay of the time auto-correlation of the (v, w) velocity field already presented in Fig. 3(c), which decays to almost zero values for time step displacements above 5.

However, more interesting than the length of the ties themselves are the "jumps" from one tie to another: that is to say, from being classified in one cluster at one time step they are then classified in another cluster at the following

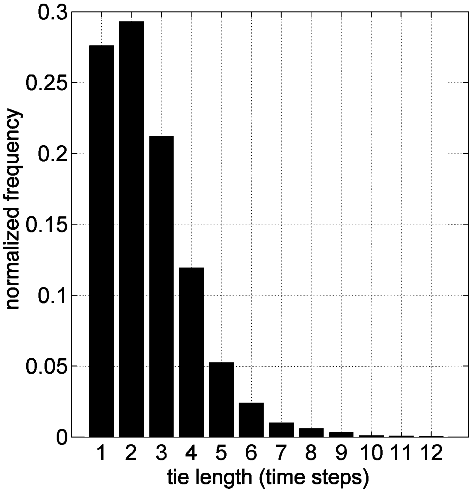

Fig. 14. Histogram of tie lengths.

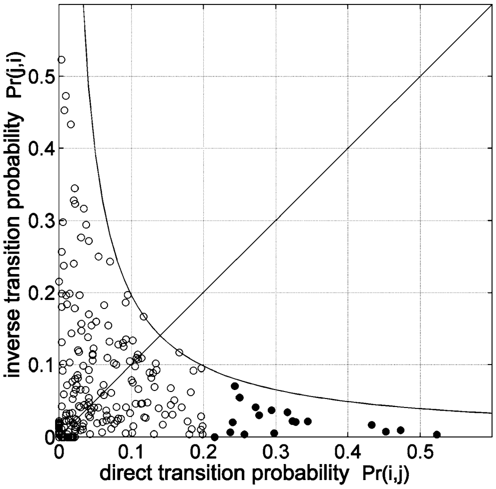

Fig. 15. Direct and inverse transition probabilities.

time step. The first jump that we observe in the short series given above as an example, is K to P, which is followed by a P-to-B jump, and so on. The last jump observed here is C to I. There are $17^2 - 17 = 272$ possible jumps, from one cluster to any other filling a matrix $Pr(i, j)$ with the probability of the (v, w) velocity field being classified into cluster V_i at time step t and into cluster V_j at time step $t + 1$. If the jumps between clusters occur randomly–that is, they are all equally probable–then all the terms in $Pr(i, j)$ should randomly fluctuate around the same value $(1/(17 - 1) \approx 6.25\%)$ and, in particular $Pr(K, P) \approx Pr(P, K)$, which is the probability of observing a K-to-P jump as above, should be of the same order of magnitude as the probability of observing a P-to-K jump.

The probability of each transition occurring was actually computed directly by counting the number of times that each transition occurred within the time series analysed. Instead of a uniform distribution of probabilities it was found that there are only a few jumps with a high probability of occurrence. Moreover, when a given jump occurs frequently, say from cluster K to cluster P, then the inverse jump, from P to K, tends to occur rarely. These results are displayed in Fig. 15, where the probability of one jump $(Pr(i, j))$ is plotted on the horizontal axis against the probability of the inverse jump $(Pr(j, i))$ on the vertical axis. Note that the plot in Fig. 15 is symmetric because both the points $(Pr(i, j), Pr(j, i))$ and $(Pr(j, i), Pr(i, j))$ are plotted. Therefore, the analysis of the results in Fig. 15 is limited to the points below the straight line $y = x$.

The first thing that can be observed in Fig. 15 is that the elements of $Pr(i, j)$ sometimes attain values as high as 0.5, but are often close to zero. Therefore, it seems that the jumps do not follow a uniform distribution. In fact only a few transitions have probabilities higher than 20%. These data points have been marked with filled circles on the horizontal axis. In addition, high values of $Pr(i, j)$ imply low values of $Pr(j, i)$, as supported by the fact that the data points with $Pr(i, j) > 0.2$ (filled circles) have $Pr(j, i)$ values well below 0.1. Only when $Pr(i, j)$ is small can $Pr(j, i)$ be of the same order.

The conclusion that can be drawn from the above observations is that some jumps between clusters are preferred, and also that they are oriented: that is to say, once it is known that a particular instantaneous (v, w) velocity field is classified into one cluster, say V_5, then it can be assessed with a probability of 0.52 that after some time steps it will fall into cluster V_8 (please see Fig. 16). And, if an instantaneous (v, w) velocity field is classified into cluster V_8, then it is fairly improbable that it will subsequently fall back into cluster V_5. This is the rightmost point in Fig. 15.

The above findings make it possible to order the clusters in the time direction. They are renumbered from 1 to 17 so that their natural time ordering is consistent with their increasing index. Fig. 16 presents the result of this ordering, which will be described in greater detail below, while Table 1 contains the matrix $Pr(i, j)$ with the most important probabilities of the jumps from cluster i to cluster j. Note that since the original order of the clusters obtained from FC is arbitrary, the original $Pr(i, j)$ matrix does not have the diagonal dominant structure of Table 1: instead it appears as random permutations of rows (and columns) of Table 1. This is why both Table 1 and Fig. 16 have not been presented until this point when they can both be presented in a more consistent way. In Fig. 16, the boxes contain the cluster index, and also the absolute probability of each centroid (in fact, these values have previously been plotted in Fig. 8).

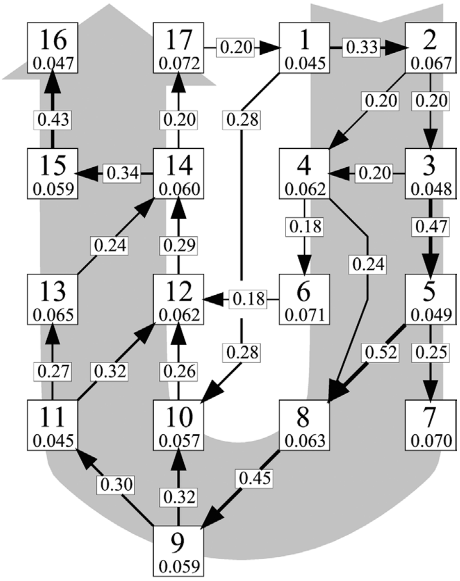

Fig. 16. Schematic representation of the most probable transitions, given arrows between pairs of clusters. The probability of these transitions is indicated in legends superimposed on the corresponding arrows. The clusters are identified with ordinal numbers (1–17) in each square box, together with the probability of each cluster (in smaller font).

Table 1
Probabilities of selected jumps. Level-1 jumps in dark cells, Level-2 jumps in grey cells

	To cluster **j**																
From cluster **i**	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1		0.33								0.28					0.01	0.11	
2			0.20	0.20													
3				0.20	0.47												
4						0.18		0.24									
5							0.25	0.52								0.05	
6												0.18					
7																0.02	
8									0.46								
9										0.32	0.30					0.01	
10												0.26				0.01	
11												0.32	0.27				
12														0.29			
13														0.24	0.34		
14																	0.20
15	0.10															0.43	
16	0.11				0.12		0.24		0.18	0.13					0.02		
17	**0.20**																

The arrows from one box to another point in the direction of the observed jumps, while the label over each arrow shows the probability of this jump. For example, we can see in Fig. 16 that the probability of randomly picking up an instantaneous (v, w) velocity field that is classified into cluster 5 is 4.9%. So, the probability that after some time steps (depending on the length of the tie) the velocity field will jump into cluster 7 is 25%, and the probability of jumping

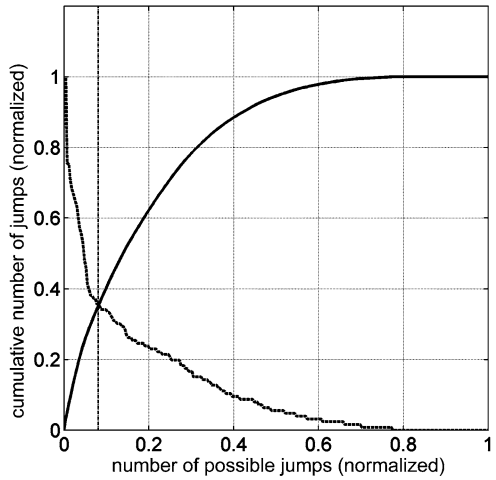

Fig. 17. Cumulative number of jumps (—-) and normalised frequency of jumps (— —), plotted in decreasing order of frequency.

into cluster 8 is 52%. In this latter case, there is a probability of 45% that the next jump observed is into cluster 9, and so on.

The general layout of Fig. 16 is rather arbitrary: we could mirror it left to right or upside down. What is less arbitrary is the information that we have included. To construct this figure we followed a simple procedure. We started by examining the matrix $Pr(i, j)$, which contains the conditional probability of each jump, and extracting the "strongest" jumps: that is to say, the $Pr(i, j)$ that were both maximum in their row and also in their column. We called this a level-1 search. Then we repeated the procedure but skipped over the values used in the previous search, in what we called a level-2 search. The lists obtained from these two searches are marked in Table 1 with dark and grey boxes, respectively, and contain the whole set of clusters, except cluster number 6. Therefore we decided to include this cluster arbitrarily in Fig. 16, and we also plotted the most probable jump to and from cluster 6 (see Table 1). We also tried to generate other plots equivalent to the one already displayed in Fig. 16, including "less strong" jumps (that is, level-3 jumps, level-4 jumps, etc). Two things happened. First, the jumps in these lists had smaller probabilities, below 0.2, and they were less oriented: that is to say, still $Pr(i, j) > Pr(j, i)$ but no longer $Pr(i, j) \gg Pr(j, i)$. Both effects can be easily observed in Fig. 15. But the most important thing was that it was impossible to keep Fig. 16 relatively simple because many cross-jumps started to appear. Still, when only level-1 and level-2 jumps were plotted, the sum of the absolute probabilities of all the jumps in Fig. 16 amounted to more than 35% of the jumps recorded in the time series. This means that if we monitored the flow for a while, one of every three jumps that we observed is in Fig. 16. Therefore, we can conclude that Fig. 16 displays the "essential dynamics" of the large-scale secondary flow motions in a square channel.

Fig. 17 plots, in decreasing order of probability, the cumulative portion of transitions as the number of possible jumps retained is increased. A vertical limit is plotted for the number of jumps retained in Fig. 16, which corresponds to 8.5% of the possible jumps and, as noted above, accounts for more than 35% of the jumps occurring in the time series. Also the decreasing normalised frequency of the jumps is plotted (please note that the fact that the three lines almost cross at the same point is a coincidence).

Fig. 17 reinforces the fact, already discussed in relation to Fig. 15, that there are a limited number of preferred or most frequent jumps, and a considerable number of jumps occur much less frequently. The most frequent jumps are the most significant for our analysis. Retaining more, but less frequent, jumps would add little significant information and increase the complexity of the analysis.

At this point the complete set of centroids obtained from the fuzzy cluster analysis of the (v, w) data can be plotted with a layout that is consistent with Fig. 16. The results can be seen in Fig. 18. In order to increase our understanding of the large-scale organisation of the flow, each centroid that was originally only the ensemble average of the data from the lower octant, has now been expanded with more points from the rest of the channel. In addition, contours of ω_x vorticity have been added to help identify the main vortices. A detailed analysis of Fig. 18 explains why the plot of the complete set of ensemble averages has been delayed until this point. Many pairs of averages linked by a

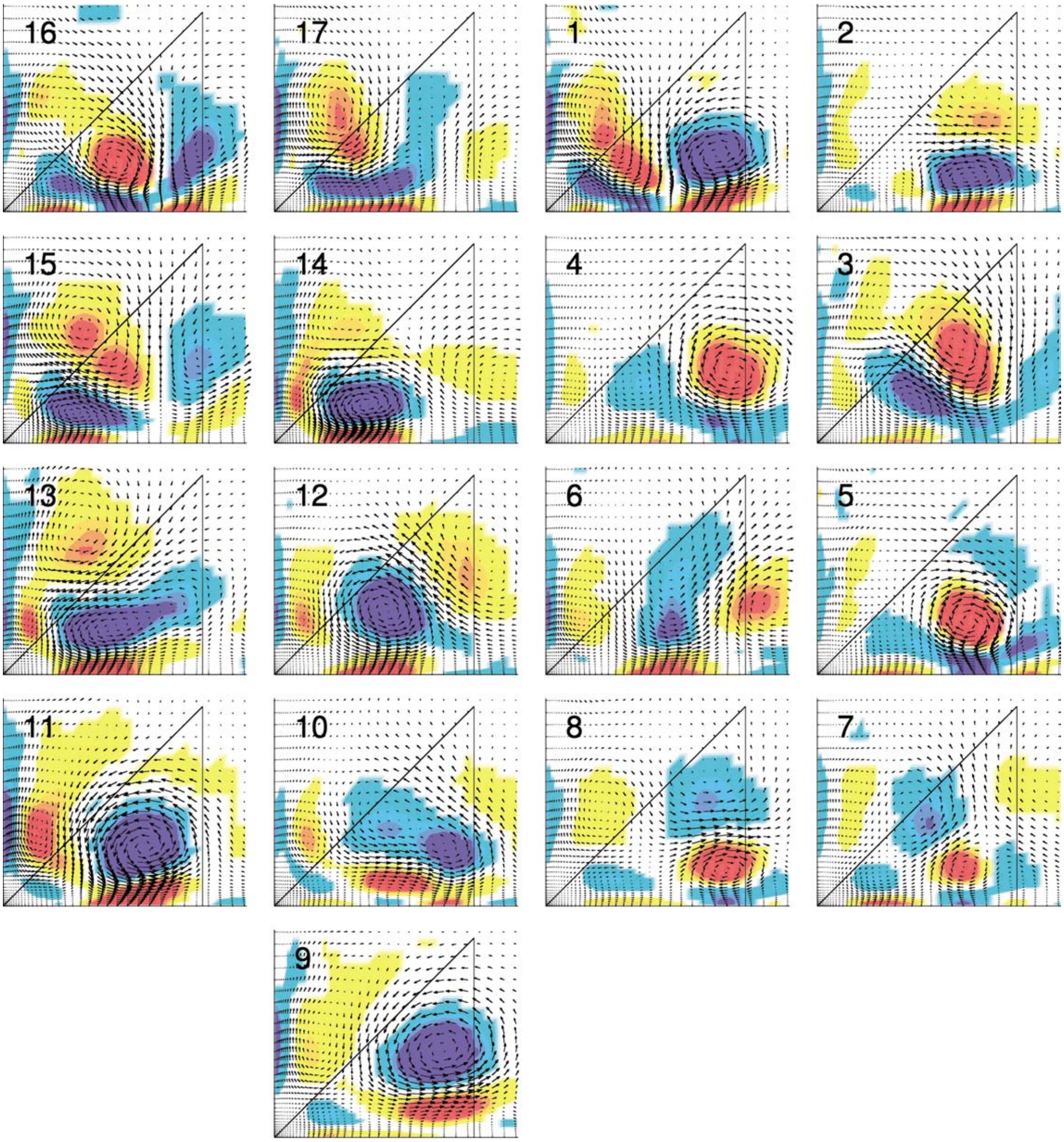

Fig. 18. Sequence of (v, w) centroids with vorticity background.

jump appear every time an eddy travels from one (z, y) position to another. This is clearly observed in the sequence 3-5-8, for example, and also in the longer sequences 9-11-12-14 and 9-11-13-14. These sequences are particularly interesting because when a large-scale vortex approaches the wall it generates a layer of sign-inverted vorticity that is no longer associated to large-scale circulation: rather it is wall-generated (boundary-layer) vorticity. Some of the conditionally sampled ensemble averages presented in Fig. 16 bear some similarity to the unconditional time mean. We can put centroids #9, #10, #11, #12, #13 and #14 in this list. Some of them, however, like centroids #4 and #5, have sign-opposed vorticity. Finally, another subset including centroids #15, #16, #17, #1 displays the traces of two counter-rotating vortices approaching the wall and, thus, lying over sign-inverted beds of vorticity. This is clearly observed in centroid #1.

Other sequences are less easy to interpret. For example, while it is clear that centroid #16 can follow #15, it is surprising that Fig. 16 shows no link between centroids #15 or #16 and centroid #1. We have to investigate the lower values of $Pr(i, j)$ to find a relationship between them. The values involving clusters #1, #15 and #16 have also

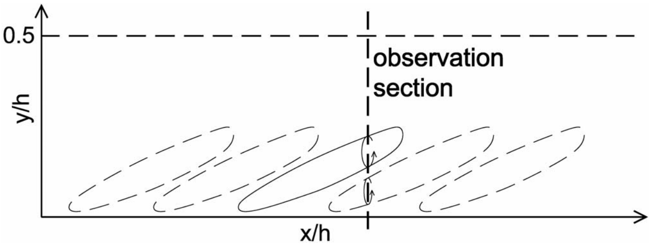

Fig. 19. Sketch of a vortical structure advected through the observation section.

been included in Table 1, which clearly shows the relationship (already displayed in Fig. 16) between #15 and #16: the probability of a 15-to-16 jump is high (43%) while the probability of a 16-to-15 jump is very low (only 1.6%). Centroid #1 has a non-symmetric relationship with #15 and #16. While 15-to-1, 16-to-1 and 1-to-16 jumps are equally probable (about 10%), the 1-to-15 jump is seldom observed (0.6%). Fig. 16 provides the explanation for these values. Centroid #1 always comes after centroids #15 and #16, and the joint probability of observing any of these jumps is 21% $(0.101 + 0.109 = 0.210)$. However, centroids #1 and #16 are too similar, and sometimes the clustering procedure misclassify an instantaneous (v, w) field and jumps back from 1-to-16.

Going further into Fig. 16's ability to represent the essential dynamics of the flow, we should point out that centroid #16 seems to be a dead end. However, a closer look at the values of $Pr(i, j)$ reveal that this is not true. Table 1 also shows the $Pr(16, j) > 0.1$ and also the corresponding $Pr(i, 16)$ values. Except for the $Pr(16, 1) \approx Pr(1, 16)$ case, discussed in the paragraph above, in all the other cases $Pr(16, j) \gg Pr(i, 16)$: that is to say, there continues to be a time directionality in the jumps. The reason why centroid #16 seems to be a dead end is clear. When the flow is classified in cluster #16 there are several possible jumps that follow this state, the most important of which is a 16-to-7 jump. This has not been included in Fig. 16 because the most important jump to cluster #7 is still from cluster #5, with probability of 25%, slightly over the 24% probability of the 16-to-7 jump.

The flow dynamics observed in Fig. 18 can be interpreted by considering the results obtained by Jeong et al. [22] in a turbulent channel, which revealed the existence of streamwise elongated vortices that originate at the wall and extend downstream and away from the wall.

Fig. 19 shows this structure in several downstream positions, as it is convected through the observation section where the data analysed here has been sampled. The intersection between the moving structure and the fixed observation section results in a vortex that approaches the wall as the structure is convected downstream, as can be seen in Fig. 18, in sequences 3-4-5-7-8 and 9-11-12-13-14-15.

This interpretation also gives further insight into the rapid loss of correlation undergone by the cross stream velocity field, as opposed to the streamwise velocity (as is shown in 3(c)). The displacement of the vortical structure produces a loss of correlation when evaluated at a given fixed cross section, since the intersection between the structure and the observation section moves and evolves. This issue is further illustrated by Fig. 20, which reproduces two instantaneous velocity frames, which follow one another within 3 time steps. They have been classified into clusters 9 and 11. While they can be interpreted as depicting the evolution of a single vortical structure like the one in Fig. 19, they have a relatively poor correlation of only 0.24.

We tried to verify this interpretation of the flow dynamics independently, without using POD and fuzzy clustering, or the λ_2 criterion used by Jeong et al. [22], and to make a quantitative estimation of the inclination angle of the structures. To this end, we computed unconditional, point to point correlation coefficients between selected locations at the wall and inner channel points. Seven points were selected along the wall at 0.05 $z/2h$ intervals. At the wall locations the streamwise component of the wall shear stress was measured, while at the inner points the selected magnitude was the streamwise vorticity, computed from the cross stream velocity components (v, w). The heights of the correlation peaks and the time delay were recorded. Finally, for each wall location, and at every wall distance, the maximum correlation peak was found. The time displacement for each of these maximum peaks was transformed into a distance using the mean convection velocity. The resulting correlation traces are plotted in Fig. 21, in normalised wall units:

$$y+ = \frac{y}{2h} \cdot Re_\tau, \qquad x+ = \frac{x}{2h} \cdot Re_\tau. \tag{10}$$

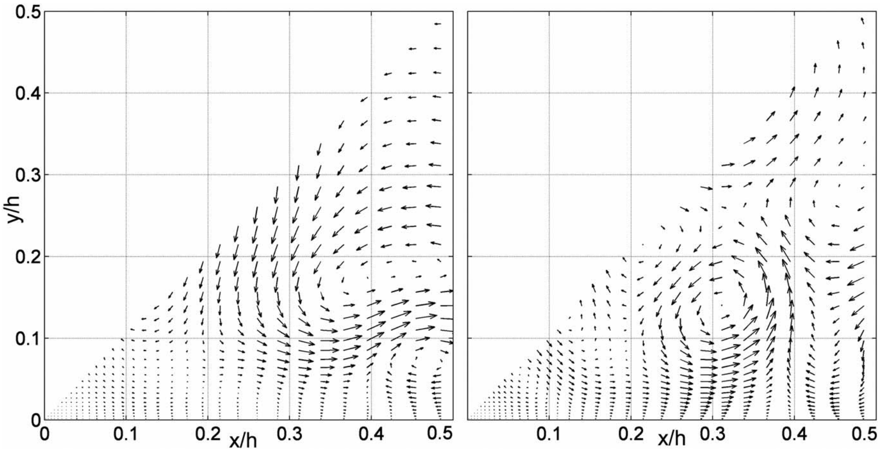

Fig. 20. Two instantaneous velocity frames, corresponding to cluster means #9 and #11, respectively. The time elapsed between both frames is 3 time steps.

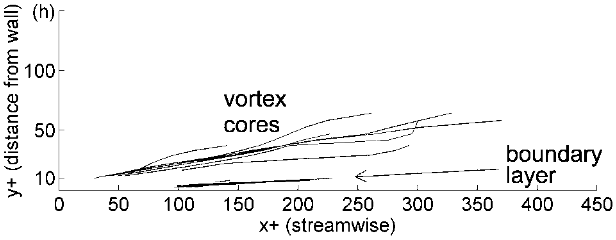

Fig. 21. Traces of the correlation peaks between streamwise vorticity and wall shear stress in the (x, y) plane.

Several aspects of this plot are worth mentioning. Firstly, two regions of collapsing traces can be seen. The main region is outside the viscous sub-layer, extending vertically from $y+ = 10$ to $y+ = 60$, and horizontally from $x+ = 50$ to $x+ = 350$. A smaller region is fully inside the viscous sub-layer, below $y+ = 10$ and from $x+ = 100$ to $x+ = 225$.

The main region is associated with the displacement of the vortical structures originated from the wall event, fixed at $x+ = 0$ in the plot. The mean inclination of these traces, in this region, is approximately 9°, in agreement with the observations of Jeong et al. [22]. The smaller region is a vorticity layer, of the opposite sign, that is forced by the structure onto the boundary layer as it moves away from the wall. These layers can be clearly seen in Fig. 18, as for example in clusters 5 and 11 which contain counter rotating vortices.

Thus, the existence of certain flow structures and their evolution over time was educed from the clustering analysis and then corroborated by an independent method. This seems a reasonable way to proceed since it exploits the unique advantages of clustering analysis, which reveals the time history of a non-periodic evolving flow but prevents any misleading conclusions that might be drawn from the probabilistic argument involved in interpreting the time relations between clusters.

6. Conclusions

A conditional sampling technique was developed, which used POD and fuzzy clustering to identify coherent structures within the flow and analyse their time history. This last property is an improvement on other techniques that are insensitive to the time ordering of the instantaneous fields. This conditional sampling technique involves all of the instantaneous fields in the time series, unlike pattern recognition or other conditional sampling schemes, for example, which typically use only a small percentage of the time series. Thus this technique is suitable for studying the large-scale flow dynamics of turbulent flows.

The turbulent flow in a square channel was analysed, with particular focus on the cross-stream secondary flow motions. The (v, w) cluster means, or ensemble averages, are much stronger than the single unconditional mean, with a much improved signal to noise ratio. This last feature makes it possible to interpret these ensemble averages directly as true representatives of the instantaneous turbulent large-scale secondary motions in this channel flow. While some of the ensemble averages obtained bear some similarity to the unconditional time mean, others clearly display sign-opposed vorticity or even traces of two counter-rotating vortices that approach the wall lying over sign-inverted beds of vorticity. The main vortical structures in the secondary flow were shown to originate at the wall, and to extend downstream and towards the centre of the channel. Vorticity layers were also observed within the viscous sub layer, forced by these vortical structures as they drift away from the wall.

When the time series formed from recording the jumps between clusters is analysed, the essential dynamics of the flow can be summarised in a relatively simple plot. This plot shows the probability of each cluster or "state" of the flow, as well as the probability of a transition to a new cluster. This makes it possible to further explore the contribution of each of the large-scale motions to the turbulent transport of momentum or heat transfer.

Acknowledgements

This work was financially supported by grants 2002 SGR00189 from AGAUR Catalan government, DPI2003-06725-C02-01 from DGI, Ministerio de Ciencia y Tecnología y Fondos FEDER, and a pre doctoral grant from the URV that was assigned to Gabriel Usera.

References

[1] A. Huser, S. Biringen, Direct numerical simulation of turbulent flow in a square duct, J. Fluid Mech. 257 (1993) 65–95.
[2] M. Piller, E. Nobile, Direct Numerical simulation of turbulent heat transfer in a square duct, Internat. J. Numer. Methods Heat Fluid Flow 12 (6) (2002) 658–686.
[3] M. Salinas Vázquez, O. Metais, Large-eddy simulation of the turbulent flow through a heated square duct, J. Fluid Mech. 453 (2002) 201–238.
[4] S. Gavrilakis, Numerical simulation of low-Reynolds-number turbulent flow through a straight square duct, J. Fluid Mech. 244 (1992) 101–129.
[5] B.K. Madabhushi, S.P. Vanka, Large eddy simulation of turbulence-driven secondary flow in a square duct, Phys. Fluids A 3 (11) (1991) 2734–2745.
[6] J. Pallares, L. Davidson, Large-eddy simulation of stationary and rotating channel and duct flows, in: Proceedings Svenska Mekanikdagar, Sweden, June 7–9, 1999, p. 4.3.
[7] H. Xu, A. Pollard, Large eddy simulation of turbulent flow in a square annular duct, Phys. Fluids 13 (11) (2001) 3321–3337.
[8] R.S. Reichert, F.F. Hatay, S. Biringen, A. Huser, Proper orthogonal decomposition applied to turbulent flow in a square duct, Phys. Fluids 6 (9) (1994) 3086–3092.
[9] G.A. Kopp, J.A. Ferré, F. Giralt, The use of pattern recognition and proper orthogonal decomposition in identifying the structure of fully-developed free turbulence, J. Fluids Engrg. 119 (1997) 289–296.
[10] G. Berkooz, P. Holmes, J.L. Lumeley, The proper orthogonal decomposition in the analysis of turbulent flows, Ann. Rev. Fluid Mech. 25 (1993) 539–575.
[11] P.B.V. Johansson, W.K. George, Proper orthogonal decomposition of an axisymmetric turbulent wake behind a disk, Phys. Fluids 14 (7) (2002) 2508–2514.
[12] S. Garnard, W.K. George, D. Jung, S. Woodwrad, Application of a "slice" proper orthogonal decomposition to the far field of an axisymmetric turbulent jet, Phys. Fluids 14 (7) (2002) 2515–2522.
[13] M. Farge, K. Schneider, G. Pellegrino, A.A. Wray, R.S. Rogallo, Coherent vortex extraction in three-dimensional homogeneous turbulence: Comparison between CVS-wavelet and POD-Fourier decompositions, Phys. Fluids 15 (10) (2003) 2886–2896.
[14] J.A. Ferré, F. Giralt, Pattern-recognition analysis of the velocity field in plane turbulent wakes, J. Fluid Mech. 198 (1989) 27–64.
[15] J.A. Ferré, J.C. Mumford, A.M. Savill, F. Giralt, Three-dimensional large eddy motions and fine-scale activity in a plane turbulent wake, J. Fluid Mech. 210 (1990) 371–414.
[16] A. Vernet, G.A. Kopp, Classification of turbulent flow patterns with fuzzy clustering, Engrg. Appl. Artificial Intelligence 15 (3–4) (2002) 315–326.
[17] J. Pallares, L. Davidson, Large-eddy simulations of turbulent heat transfer in stationary and rotating square ducts, Phys. Fluids 14 (8) (2002) 2804–2816.
[18] S. Gavrilakis, Turbulent-field modes in the vicinity of the walls of a square duct, Eur. J. Mech. B Fluids 16 (1997) 121–139.
[19] J. Pallares, L. Davidson, Large-eddy simulation of turbulent flows in stationary and rotating channels and in a stationary duct", Rep 00/03. Department of Thermo and Fluid Dynamics, Chalmers University of Technology Gothenburg, Sweden, 2000 (available online at http://www.tfd.chalmers.se/~lada/postscript_files/pallares_tfd00-03.pdf, as of 08/Feb/2005).
[20] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, New York, 1981.
[21] X.L. Xie, G. Beni, A validity measure for fuzzy clustering, IEEE Trans. Pattern Anal. Machine Intell. PAMI-13 (8) (1991) 841–847.
[22] J. Jeong, F. Hussain, W. Schoppa, J. Kim, Coherent structures near the wall in a turbulent channel flow, J. Fluid Mech. 332 (1997) 185–214.